

**Penerapan *Data Mining* Pemilihan Siswa Kelas Unggulan  
dengan Metode *K-Means Clustering* di SMP N 02 Tasikmadu**



**Disusun sebagai salah satu syarat menyelesaikan Program Studi Strata I pada  
Jurusan Informatika Fakultas Komunikasi dan Informatika**

**Oleh:**

**TEGUH WIBOWO**

**L 200144019**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS KOMUNIKASI DAN INFORMATIKA  
UNIVERSITAS MUHAMMADIYAH SURAKARTA  
2018**

**HALAMAN PERSETUJUAN**

**Penerapan *Data Mining* Pemilihan Siswa Kelas Unggulan  
dengan Metode *K-Means Clustering* di SMP N 02 Tasikmadu**

**PUBLIKASI ILMIAH**

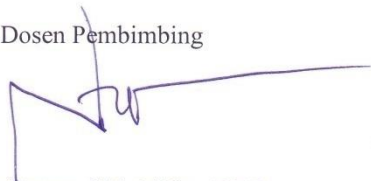
oleh:

**TEGUH WIBOWO**

**L 200144019**

Telah diperiksa dan disetujui untuk diuji oleh:

Dosen Pembimbing



**Nurgiyatna, S.T., M.Sc., Ph.D.**

**NIK.881**

HALAMAN PENGESAHAN

**Penerapan *Data Mining* Pemilihan Siswa Kelas Unggulan  
dengan Metode *K-Means Clustering* di SMP N 02 Tasikmadu**

OLEH

**TEGUH WIBOWO**

L 200144019

Telah dipertahankan di depan Dewan Penguji

Fakultas Komunikasi dan Informatika

Universitas Muhammadiyah Surakarta

Pada hari Rabu, 31 Januari 2018

dan dinyatakan telah memenuhi syarat

Dewan Penguji :

1. Nurgiytna, S.T.,M.Sc, Ph.D.

(Ketua Dewan Penguji)

2. Dr.,Ir. Bana Handaga, M.T.

(Anggota I Dewan Penguji)

3. Aris Rakhmadi, S.T., M.Eng.

(Anggota II Dewan Penguji)

(.....)

(.....)

(.....)

Dekan,

Fakultas Komunikasi dan Informatika



Nurgiytna, S.T.,M.Sc, Ph.D.

NIK. 881

Ketua Program Studi Informatika



Dr. Heru Supriyono, S.T., M.Sc.

NIK. 970

## PERNYATAAN

Dengan ini saya menyatakan bahwa dalam naskah publikasi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan orang lain, kecuali secara tertulis diacu dalam naskah dan disebutkan dalam daftar pustaka.

Apabila kelak terbukti ada ketidakbenaran dalam pernyataan saya di atas, maka akan saya pertanggungjawabkan sepenuhnya.

Surakarta, 31 Januari ..... 2018

Penulis



TEGUH WIBOWO

L 200144019



**UNIVERSITAS MUHAMMADIYAH SURAKARTA  
FAKULTAS KOMUNIKASI DAN INFORMATIKA  
PROGRAM STUDI INFORMATIKA**

Jl. A Yani Tromol Pos 1 Pabelan Kartasura Telp. (0271)717417, 719483 Fax (0271) 714448  
Surakarta 57102 Indonesia. Web: <http://informatika.ums.ac.id>. Email: [informatika@ums.ac.id](mailto:informatika@ums.ac.id)

**SURAT KETERANGAN LULUS PLAGIASI**

162/A.3-II.3/inf-FKI/IV/2018

Assalamu'alaikum Wr. Wb

Biro Skripsi Program Studi Informatika menerangkan bahwa :

Nama : TEGUH WIBOWO  
NIM : L200144019  
Judul : Penerapan *Data Mining* Pemilihan Siswa Kelas Unggulan dengan  
Metode *K-Means Clustering* di SMP N 02 Tasikmadu  
Program Studi : Informatika  
Status : **Lulus**

Adalah benar-benar sudah lulus pengecekan plagiasi dari Naskah Publikasi Skripsi, dengan menggunakan aplikasi Turnitin.

Demikian surat keterangan ini dibuat agar dipergunakan sebagaimana mestinya.

Wassalamu'alaikum Wr. Wb

Surakarta, 10 April 2018

Biro Skripsi Informatika

**Ihsan Cahyo Utomo, S.Kom., M.Kom.**



UNIVERSITAS MUHAMMADIYAH SURAKARTA  
FAKULTAS KOMUNIKASI DAN INFORMATIKA  
PROGRAM STUDI INFORMATIKA

Jl. A Yani Tromol Pos 1 Pabelan Kartasura Telp. (0271)717417, 719483 Fax (0271) 714448  
Surakarta 57102 Indonesia. Web: <http://informatika.ums.ac.id>. Email: [informatika@ums.ac.id](mailto:informatika@ums.ac.id)

Feedback Studio - Google Chrome  
Secure | [https://ev.turnitin.com/app/carta/en\\_us/?o=943944341&s=1&lang=en-us&u=1057550080](https://ev.turnitin.com/app/carta/en_us/?o=943944341&s=1&lang=en-us&u=1057550080)

feedback studio | Penerapan Data Mining Pemilihan Siswa Kelas Unggulan dengan Metode K-Means Clustering di SMP N 02 Tasikmadu

**Penerapan Data Mining Pemilihan Siswa Kelas Unggulan  
dengan Metode K-Means Clustering di SMP N 02 Tasikmadu**

**Abstrak**

SMP N 02 Tasikmadu merupakan sekolah yang berada di kota Karanganyar. Sekolah ini memiliki sistem pembelajaran yang baik yaitu program kelas siswa unggulan. SMP N 02 Tasikmadu mengalami peningkatan penerimaan pendaftar siswa baru dalam setiap tahunnya. Banyaknya siswa yang mendaftar maka dalam pemilihan siswa kelas unggulanpun mengakibatkan kesulitan dalam menentukan kelas unggulan yang sesuai dengan bakat kemampuan siswa. Oleh sebab itu penerapan *data mining* ini dilakukan untuk membantu keputusan dalam memilih pengelompokan kelas unggulan dengan acuan nilai-nilai yang dimiliki siswa dengan menggunakan metode *k-means clustering*. Sebagai penerapan metode *clustering* untuk data perhitungan algoritma *k-means* yang digunakan adalah nilai rapor siswa dan masing-masing data nilai digunakan sebagai atribut. Atribut-atribut yang dipilih diterapkan menggunakan metode *k-means clustering* untuk menghasilkan 5 *cluster* yang diambil 3 *cluster* untuk kelas unggulan. Hasil dari penelitian ini menunjukkan bahwa algoritma *k-means* mampu menghasilkan pemilihan dan pembagian kelas unggulan sesuai nilai kemampuan siswa.

Page: 1 of 15 | Word Count: 2986 | Text-only Report

**Match Overview**

29%

1	<a href="http://eprints.ums.ac.id">eprints.ums.ac.id</a> Internet Source	6%
2	<a href="http://media.neliti.com">media.neliti.com</a> Internet Source	2%
3	<a href="http://pelita-informatika.com">pelita-informatika.com</a> Internet Source	2%
4	Submitted to Universita... Student Paper	2%
5	Submitted to IGroup Student Paper	1%
6	<a href="http://publikasilmiah.ums.ac...">publikasilmiah.ums.ac...</a> Internet Source	1%
7	<a href="http://repository.upi.edu">repository.upi.edu</a> Internet Source	1%

**Penerapan *Data Mining* Pemilihan Siswa Kelas Unggulan  
dengan Metode *K-Means Clustering* di SMP N 02 Tasikmadu**

**Abstrak**

SMP N 02 Tasikmadu merupakan sekolah yang berada di kota Karanganyar. Sekolah ini memiliki sistem pembelajaran yang baik yaitu program kelas siswa unggulan. SMP N 02 Tasikmadu mengalami peningkatan penerimaan pendaftar siswa baru dalam setiap tahunnya. Banyaknya siswa yang mendaftar maka dalam pemilihan siswa kelas unggulanpun mengakibatkan kesulitan dalam menentukan kelas unggulan yang sesuai dengan bakat kemampuan siswa. Oleh sebab itu penerapan *data mining* ini dilakukan untuk membantu keputusan dalam memilih pengelompokkan kelas unggulan dengan acuan nilai-nilai yang dimiliki siswa dengan menggunakan metode *k-means clustering*. Sebagai penerapan metode *clustering* untuk data perhitungan algoritma *k-means* yang digunakan adalah nilai raport siswa dan masing-masing data nilai digunakan sebagai atribut. Atribut-atribut yang dipilih diterapkan menggunakan metode *k-means clustering* untuk menghasilkan 5 *cluster* yang diambil 3 *cluster* untuk kelas unggulan. Hasil dari penelitian ini menunjukkan bahwa algoritma *k-means* mampu menghasilkan pemilihan dan pembagian kelas unggulan sesuai nilai kemampuan siswa.

**Kata Kunci :** *Data mining, K-Means clustering*, pembagian kelas.

**Abstract**

Tasikmadu 02 junior high school is a school located in Karanganyar city. This school has a good learning system that is a superior student class program. Tasikmadu 02 junior high school has increased enrollment of new students every year. The number of students who register then in the selection of superior studentspace resulted in the difficulty in determining the superior classroom in accordance with the ability of students.. Therefore the application of data mining is done to help the distribution of superior classes with reference values that students have using the method of k-means clustering. As the application of the clustering method for k-means algorithm calculation data used is the student's raport score and each value data is used as an attribute. The selected attributes are applied using k-means clustering method to produce 5 clusters taken 3 clusters for the superior class. The results of this study indicate that the k-means algorithm is able to produce selection and distribution of superior class according to the student's ability.

**Keywords :** *Data mining, K-Means clustering*, class division.

## 1. PENDAHULUAN

Sekolah Menengah Pertama (SMP) adalah jenjang pendidikan dasar pada pendidikan formal di Indonesia setelah lulus sekolah dasar atau sederajat. Sekolah menengah pertama ditempuh dalam waktu 3 tahun, mulai dari kelas 7 sampai kelas 9. Pada tahun ajaran 1994/1995 hingga 2003/2004, sekolah ini pernah disebut sekolah lanjutan tingkat pertama (SLTP). SMP N 02 Tasikmadu merupakan sekolah menengah pertama yang beralamat di Mulyorejo Kalijirak Tasikmadu Karanganyar. SMP ini memiliki sistem pembelajaran yang baik dan program siswa kelas unggulan. SMP N 02 Tasikmadu dalam penerimaan pendaftaran siswa baru meningkat setiap tahunnya dan menjadi sekolah favorit di Karanganyar.

Dengan bertambahnya pendaftar siswa baru setiap tahun maka data-data siswa yang dimiliki juga semakin bertambah banyak. Meningkatnya jumlah data siswa mengakibatkan kesulitan dalam menentukan siswa kelas unggulan dalam mengelompokkan siswa sesuai dengan nilai bakat kemampuan. Salah satu cara untuk mengatasi masalah ini adalah dengan menggunakan teknik *data mining* yang bisa digunakan untuk pengolahan data menjadi sumber informasi strategis. *Data mining* dapat membantu sebuah organisasi yang memiliki data melimpah untuk memberikan informasi yang dapat mendukung pengambilan keputusan (Bhardwaj et al, 2012).

Berdasarkan permasalahan tersebut, maka dalam penelitian ini akan mengklustering siswa kelas unggulan menggunakan metode *clustering* dengan algoritma *k-means*, serta pemilihan atribut sesuai dengan kebutuhan. Kelas unggulan yang diharapkan adalah 3 kelas favorit meliputi siswa-siswa pilihan yang memiliki nilai rapot sesuai dengan standar nilai kemampuan siswa.

Tujuan penelitian ini adalah mengklustering siswa kelas unggulan serta memberikan rencana strategis bagi SMP N 02 Tasikmadu berdasar hasil penelitian. Sehingga dapat tercipta peningkatan proses belajar sesuai nilai bakat kemampuan siswa untuk mencapai hasil optimal.



## 1.1 Landasan Teori

### 1.1.1 Data Mining

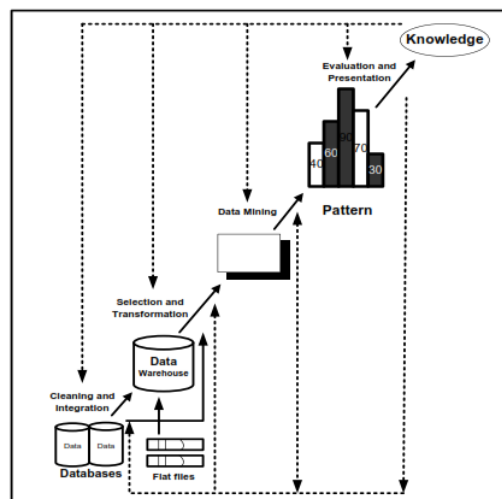
*Data mining* adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran *computer* (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis. Definisi lain diantaranya adalah pembelajaran berbasis induksi (*induction-based learning*) adalah proses pembentukan definisi-definisi konsep umum yang dilakukan dengan cara mengobservasi contoh-contoh spesifik dari konsep-konsep yang akan dipelajari. *Knowledge Discovery in Databases* (KDD) adalah penerapan metode pada *data mining*. Dalam konteks ini *data mining* merupakan satu langkah dari proses KDD (Abdillah, 2013).

Beberapa teknik dan sifat *data mining* adalah sebagai berikut:

- 1) *Classification*
- 2) *Clustering*
- 3) *Association Rule*
- 4) *Regression*
- 5) *Deviation Detection*

#### Tahap-Tahap *Data Mining*

*Data mining* adalah sebuah untaian proses, maka terbagi menjadi beberapa tahap. Tahapan tersebut akan bersifat interaktif, pengguna akan terlibat langsung atau dengan perantara KDD (Untari, 2014). Berikut tahapan *data mining* ditunjukkan pada gambar 1.



Gambar 1. Tahapan *data mining*

Tahapan *data mining* dibagi menjadi bagian-bagian yaitu :

- 1) Pembersihan data (*data cleaning*)
- 2) Integrasi data (*data integration*)
- 3) Seleksi data (*Data Selection*)
- 4) Transformasi data (*Data Transformation*)
- 5) Proses *mining*
- 6) Evaluasi pola (*pattern evaluation*)
- 7) Presentasi pengetahuan (*knowledge presentation*)(Sujarweni, 2014).

#### 1.1.2 *Clustering*

*Clustering* atau klasterisasi merupakan suatu metode atau teknik pengelompokan data. *Clustering* berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam *clustering*. *Clustering* tidak mencoba untuk melakukan klasifikasi ataupun memprediksi nilai dari variabel target. Akan tetapi, proses ini mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (homogen). Teknik *clustering* banyak diterapkan dalam berbagai bidang. Misalnya dalam bidang medis, *clustering* dapat digunakan untuk mengelompokkan jenis-jenis penyakit berdasarkan karakteristik dan gejala-gejala yang dialami pasien (Prasetyo, 2012).

#### 1.1.3 Algoritma *K-Means*

*K-means* merupakan salah satu metode data klastering non-hirarki untuk mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster*. Metode ini mempartisi data ke dalam *cluster* sehingga data yang memiliki karakteristik sama dikelompokkan ke dalam satu *cluster* dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam *cluster* lain(Prasetyo, 2012). Dalam penerapannya, *k-means* memisahkan data dengan melakukan perulangan secara terus-menerus sampai tidak ada perubahan data dalam setiap segmentasi. Langkah-langkah dalam metode *k-means* adalah sebagai berikut:

- 1) Tentukan nilai *k* sebagai jumlah klaster yang ingin dibentuk.
- 2) Bangkitkan *k centroid* (titik pusat *cluster*) awal.
- 3) Hitung jarak setiap data ke masing-masing *centroid* menggunakan rumus *Manhattan Distance* seperti pada persamaan (1).

$$D(X,Y) = \sum |X_i - Y_i| \dots\dots\dots (1)$$

Keterangan:

$D(X,Y)$  = jarak objek antara  $X_i$  dan  $Y_i$

$X_i$  = Koordinat dari objek  $X_i$  pada dimensi  $i$

$Y_i$  = Koordinat dari objek  $Y_i$  pada dimensi  $i$

- 4) Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan *centroid*nya.
- 5) Tentukan *centroid* baru dengan cara menghitung nilai rata-rata dari data-data yang ada pada *centroid* yang sama dengan rumus seperti persamaan (2). Dimana  $n$  adalah banyaknya dokumen dalam *cluster*  $i$  dan  $x$  adalah data yang akan dihitung.

$$c_i = \left( \frac{\sum x}{n} \right) \dots\dots\dots (2)$$

- 6) Kembali ke langkah 3 jika posisi *centroid* baru dengan *centroid* lama tidak sama.

#### 1.1.4 RapidMiner

RapidMiner merupakan *software tool Open Source* untuk *data mining*. RapidMiner dioperasikan pada sebuah lingkungan untuk *machine learning*, *data mining*, *text mining* dan *predictive analytics* (Romi, 2012).

## 2. METODE

### 2.1 Metode Pengumpulan Data

Metode pengumpulan data yang digunakan dalam penelitian ini menggunakan metode Studi Pustaka. Metode ini metode yang digunakan untuk mendapatkan data yang dibutuhkan mencari informasi mengenai pengolahan data baik dari buku-buku, laporan penelitian maupun jurnal dari internet. Sumber data yang digunakan dalam penelitian ini yaitu:

#### a. Data Primer

Data primer diperoleh secara langsung dari sumber data yang berhubungan dengan penelitian yang dilakukan yaitu data nilai siswa SMP N 02 Tasikmadu.

#### b. Data Sekunder

Data sekunder sebagai pelengkap data primer yang diperoleh dari buku-buku, jurnal ilmiah, publikasi, laporan penelitian dan internet untuk menunjang teori. Pada metode ini kegiatan yang dilakukan adalah mempelajari, mencari dan mengumpulkan data-data yang berhubungan dengan

penentuan siswa kelas unggulan SMP N 02 Tasikmadu. Data yang telah diperoleh kemudian akan diolah menggunakan metode *clustering* algoritma *k-means* dengan mengambil nilai dari setiap atribut pada data untuk menentukan kelas siswa unggulan.

## **2.2 Metode Analisis Data**

Analisis data merupakan suatu proses mencari dan menyusun secara sistematis data yang telah diperoleh dari wawancara, observasi, dan lain sebagainya, dengan cara mengorganisasikan data tersebut ke dalam kategori, memilih mana yang penting dan mana yang akan dipelajari dan kemudian membuat kesimpulan agar dapat dipahami diri sendiri maupun orang lain.

Dalam penulisan penelitian ini menggunakan analisis data kuantitatif yaitu analisis yang dilakukan dengan mengelompokkan data untuk mencari suatu pola dari hal-hal yang dipelajari dalam konsep yang ada didalam sumber.

## **2.3 Diagram Alur Penelitian**

Dalam penelitian ini langkah-langkah yang digunakan untuk acuan sebagai urutan penelitian disajikan pada gambar 4. Langkah-langkah penelitian sebagai berikut :

### **1) Data Siswa**

Mulai untuk melakukan penelitian, menyiapkan dan mengumpulkan data yang dibutuhkan dan berhubungan dengan penelitian.

### **2) *Preprocessing***

Tahap kedua yaitu dengan melakukan *preprocessing* data. *Preprocessing* merupakan pengolahan awal data dan mempersiapkan data teks untuk dilakukan proses klasterisasi, yaitu dengan melakukan metode :

- a) Integrasi data
- b) Seleksi data

### **3) Proses algoritma *k-means***

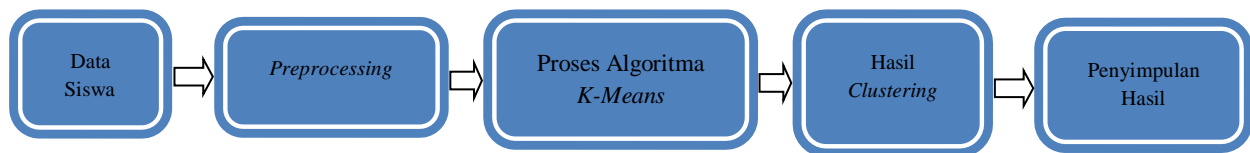
Mengimplementasikan algoritma *k-means* untuk mengklasterisasi data yang digunakan. Secara manual menggunakan program aplikasi *Microsoft Office Excel*.

### **4) Hasil *Clustering***

Memberikan hasil pengelompokan siswa kelas unggulan/favorit.

## 5) Penyimpulan Hasil

Memberikan kesimpulan atas hasil penelitian.



Gambar 2. Diagram alur penelitian

## 2.4 Data Penelitian

Data penelitian berupa nilai rapot siswa kelas 7 yang terdiri 7 kelas dan masing-masing data memiliki atribut yang digunakan dalam proses perhitungan untuk menentukan siswa kelas unggulan. Berikut atribut data siswa disajikan pada tabel 1.

Tabel 1. Atribut data siswa

Nilai Agama	Atribut yang menginformasikan nilai pelajaran agama
Nilai PKN	Atribut yang menginformasikan nilai pelajaran PKN
Nilai B. Indonesia	Atribut yang menginformasikan nilai pelajaran B. Indonesia
Nilai B. Inggris	Atribut yang menginformasikan nilai pelajaran B. Inggris
Nilai Matematika	Atribut yang menginformasikan nilai pelajaran Matematika
Nilai IPA	Atribut yang menginformasikan nilai pelajaran IPA
Nilai IPS	Atribut yang menginformasikan nilai pelajaran IPS
Nilai Seni Budaya	Atribut yang menginformasikan nilai pelajaran Seni Budaya
Nilai Pendidikan Jasmani Olahraga dan Kesehatan	Atribut yang menginformasikan nilai pelajaran Pendidikan Jasmani Olahraga dan Kesehatan
Nilai TIK	Atribut yang menginformasikan nilai pelajaran TIK
Nilai B. Jawa	Atribut yang menginformasikan nilai pelajaran B. Jawa
Nilai Ekstrakurikuler	Atribut yang menginformasikan nilai ekstrakurikuler siswa
Kepribadian	Atribut yang menginformasikan nilai kepribadian siswa
Tata Busana	Atribut yang menginformasikan nilai ketrampilan tata busana siswa

## 2.5 Seleksi Data

Pada data siswa atribut di seleksi dan dipilih untuk digunakan dalam proses *mining*. Berikut seleksi atribut seperti yang ada pada tabel 2.

Tabel 2. seleksi atribut

Atribut	v (ya) / x (tidak)
Nilai Agama	V
Nilai PKN	V
Nilai B. Indonesia	V
Nilai B. Inggris	V
Nilai Matematika	V
Nilai IPA	V
Nilai IPS	V
Nilai Seni Budaya	V
Nilai Pendidikan Jasmani Olahraga dan Kesehatan	V
Nilai TIK	X
Nilai B. Jawa	V
Nilai Ekstrakurikuler	X
Kepribadian	X
Tata Busana	V

Berdasarkan tabel 2 diatas, maka atribut-atribut yang akan digunakan dalam penelitian tersaji pada tabel 3.

Tabel 3. Atribut data penelitian

No	Atribut
1	Nilai Agama
2	Nilai PKN
3	Nilai B. Indonesia
4	Nilai B. Inggris
5	Nilai Matematika
6	Nilai IPA
7	Nilai IPS
8	Nilai S. Budaya
9	Nilai Pendidikan Jasmani Olahraga dan Kesehatan
10	Nilai B. Jawa
11	Nilai Tata Busana

Sebelum penerapan proses perhitungan *k-means clustering* pada penelitian setelah menyeleksi atribut yaitu menentukan kriteria yang digunakan untuk menyeleksi siswa. Nilai pada tabel 4 digunakan untuk menyeleksi nilai siswa untuk kelas favorit.

Tabel 4. Tabel nilai

Mata Pelajaran	Standar Nilai
P. Agama	55
Indonesia	40
Inggris	35
Matematika	40
IPA	35
IPS	40

### 3. HASIL DAN PEMBAHASAN

Data-data yang diperoleh pada tahapan pengumpulan data menghasilkan 226 data siswa dengan atribut Agama, PKn, Bahasa Indonesia, Bahasa Inggris, Matematika, IPA, IPS, Seni Budaya, Penjas, TIK, Bahasa Jawa, Tata Busana yang nantinya digunakan sebagai data perhitungan *k-means*. Berikut potongan hasil *preprocessing* data siswa yang disajikan pada tabel 5.

Tabel 5. Potongan hasil *preprocessing* data siswa

No	NIS	Nama	P. Agama	PKn	B. Indo	B. Ing	MTK	IPA	IPS	S. Bud	Penjas	B. Jawa	Tata Busana
1	4745	Adna Maksal Hafizto	74	70	72	45	48	45	73	86	60	72	58
2	4746	Aisyah Mutiara C.	82	70	86	65	56	43	68	88	68	86	69
3	4747	Akhid Rijal F.	56	84	62	40	48	48	60	58	65	76	75
4	4748	Alfian Fajar Nur P.	80	89	75	60	60	48	75	46	64	78	71
5	4749	Annisa Nur Eka Putri	88	74	78	50	48	38	63	60	66	66	69
6	4750	Annita Rahmawati	88	82	82	65	56	63	70	72	65	92	74
7	4751	Banu Nurcahyo	70	76	74	45	68	48	70	78	67	68	59
8	4752	Dewi Wulan Ningsih	90	90	74	55	64	38	55	68	61	56	71
9	4753	Dony Hartanto	62	76	63	60	36	45	78	64	62	70	58
10	4754	Dwi Pangiling	79	63	74	45	48	43	70	68	60	68	58

Data-data hasil *preprocessing* data siswa diseleksi sesuai standar nilai yang sudah ditentukan menghasilkan 159 data siswa dan akan digunakan dalam proses perhitungan *k-means clustering*. Berikut potongan hasil seleksi data siswa disajikan pada tabel 6.

Tabel 6. Potongan hasil seleksi data siswa sesuai standar nilai

NIS	Nama	P. Agama	PKn	B. Indo	B. Ing	MTK	IPA	IPS	S. Bud	Penjas	B. Jawa	Tata Busana
4745	Adna Maksimal Hafizto	74	70	72	45	48	45	73	86	60	72	58
4746	Aisyah Mutiara C.	82	70	86	65	56	43	68	88	68	86	69
4747	Akhid Rijal F.	56	84	62	40	48	48	60	58	65	76	75
4748	Alfian Fajar Nur P.	80	89	75	60	60	48	75	46	64	78	71
4750	Annita Rahmawati	88	82	82	65	56	63	70	72	65	92	74
4751	Banu Nurcahyo	70	76	74	45	68	48	70	78	67	68	59
4754	Dwi Pangling	79	63	74	45	48	43	70	68	60	68	58
4757	Juli Iskandar	83	85	84	75	70	68	93	82	62	82	70
4759	Nukke Tantina Artha	78	66	90	50	64	53	83	66	62	76	56
4761	Panji Eka Sagita	73	81	41	35	52	40	80	66	55	86	60

### 3.1 Proses Manual K-Means Clustering

Tahapan proses perhitungan secara manual yaitu:

1) Menentukan nilai k sebagai jumlah *cluster* yang ingin dibentuk.

Dalam menentukan jumlah *cluster* dari data-data yang ada akan dibuat menjadi 5 *cluster*.

2) Menentukan k *centroid* (titik pusat *cluster*) awal.

Di ambil data ke-49 sebagai pusat *cluster* ke-1 98 88 84 75 78 88 95 92 68 90 90

Di ambil data ke-187 sebagai pusat *cluster* ke-2 94 83 68 65 88 50 85 80 64 86 62

Di ambil data ke-190 sebagai pusat *cluster* ke-3 90 94 78 50 68 43 80 82 64 62 64

Di ambil data ke-161 sebagai pusat *cluster* ke-4 60 94 80 55 60 45 73 52 65 76 70

Di ambil data ke-141 sebagai pusat *cluster* ke-5 72 63 62 35 40 53 73 81 58 84 66

3) Menghitung jarak setiap data ke masing-masing *centroid*

$$D|X_1, C_1| = |X_{i1} - C_{j1}| + |X_{i2} - C_{j2}| + |X_{i3} - C_{j3}| + |X_{i4} - C_{j4}| + |X_{i5} - C_{j5}| + |X_{i6} - C_{j6}| + |X_{i7} - C_{j7}| + |X_{i8} - C_{j8}| + |X_{i9} - C_{j9}| + |X_{i10} - C_{j10}| + |X_{i11} - C_{j11}|$$

$$= |98 - 74| + |88 - 70| + |84 - 72| + |75 - 45| + |78 - 48| + |88 - 45| + |95 - 73| + |92 - 86| + |68 - 60| + |90 - 72| + |90 - 58| = 243$$

$$D|X_1, C_2| = |X_{i1} - C_{j1}| + |X_{i2} - C_{j2}| + |X_{i3} - C_{j3}| + |X_{i4} - C_{j4}| + |X_{i5} - C_{j5}| + |X_{i6} - C_{j6}| + |X_{i7} - C_{j7}| + |X_{i8} - C_{j8}| + |X_{i9} - C_{j9}| + |X_{i10} - C_{j10}| + |X_{i11} - C_{j11}|$$

$$= |94 - 74| + |83 - 70| + |68 - 72| + |65 - 45| + |88 - 48| + |50 - 45| + |85 - 73| + |80 - 86| + |64 - 60| + |86 - 72| + |62 - 58| = 142$$

$$D|X_1, C_3| = |X_{i1} - C_{j1}| + |X_{i2} - C_{j2}| + |X_{i3} - C_{j3}| + |X_{i4} - C_{j4}| + |X_{i5} - C_{j5}| + |X_{i6} - C_{j6}| + |X_{i7} - C_{j7}| + |X_{i8} - C_{j8}| + |X_{i9} - C_{j9}| + |X_{i10} - C_{j10}| + |X_{i11} - C_{j11}|$$

$$= |90 - 74| + |94 - 70| + |78 - 72| + |50 - 45| + |68 - 48| + |43 - 45| + |80 - 73| + |82 - 86| + |64 - 60| + |62 - 72| + |64 - 58| = 104$$

$$D|X_1, C_4| = |X_{i1} - C_{j1}| + |X_{i2} - C_{j2}| + |X_{i3} - C_{j3}| + |X_{i4} - C_{j4}| + |X_{i5} - C_{j5}| + |X_{i6} - C_{j6}| + |X_{i7} - C_{j7}| + |X_{i8} - C_{j8}| + |X_{i9} - C_{j9}| + |X_{i10} - C_{j10}| + |X_{i11} - C_{j11}|$$

$$= |60 - 74| + |94 - 70| + |80 - 72| + |55 - 45| + |60 - 48| + |45 - 45| + |73 - 73| + |52 - 86| + |65 - 60| + |76 - 62| + |70 - 58| = 123$$

$$D|X_1, C_5| = |X_{i1} - C_{j1}| + |X_{i2} - C_{j2}| + |X_{i3} - C_{j3}| + |X_{i4} - C_{j4}| + |X_{i5} - C_{j5}| + |X_{i6} - C_{j6}| + |X_{i7} - C_{j7}| + |X_{i8} - C_{j8}| + |X_{i9} - C_{j9}| + |X_{i10} - C_{j10}| + |X_{i11} - C_{j11}|$$

$$= |72 - 74| + |63 - 70| + |62 - 72| + |35 - 45| + |40 - 48| + |53 - 45| + |73 - 73| + |81 - 86| + |58 - 60| + |84 - 62| + |66 - 58| = 72$$



Berikut potongan hasil perhitungan jarak setiap data ke masing-masing *centroid* tersaji dalam tabel 7.

Tabel 7. Potongan hasil perhitungan jarak setiap data ke masing-masing *centroid*

NIS	Nama	P. Agama	PKn	B. Indo	B. Ing	MTK	IPA	IPS	S. Bud	Pengas	B. Jawa	Tata Busana	C1	C2	C3	C4	C5	Jarak Terpendek
4745	Adna Maksimal Hafizto	74	70	72	45	48	45	73	86	60	72	58	243	142	104	123	72	72
4746	Aisyah Mutiara C.	82	70	86	65	56	43	68	88	68	86	69	169	118	118	123	124	118
4747	Akhid Rijal F.	56	84	62	40	48	48	60	58	65	76	75	274	183	165	86	115	86
4748	Alfian Fajar Nur P.	80	89	75	60	60	48	75	46	64	78	71	202	123	105	50	151	50
4750	Annita Rahmawati	88	82	82	65	56	63	70	72	65	92	74	141	108	126	117	146	108

4) Mengelompokkan setiap data berdasarkan jarak terdekat antara data dengan *centroid*nya seperti pada tabel 8.

Tabel 8. Potongan hasil pengelompokkan data 1(iterasi 1)

C1	C2	C3	C4	C5
				1
	1	1		
			1	
			1	
	1			

5) Menentukan posisi *centroid* baru dengan cara menghitung nilai rata-rata dari data-data yang ada pada *centroid* yang sama. Dimana n adalah banyaknya dokumen dalam *cluster* i dan x adalah data yang akan dihitung. Hasil perhitungan untuk *centroid* baru ditunjukkan pada tabel 9.

$$C_{1x1} = (83+99+98+81+97+98+97+89+94)/10 = 83,6$$

Tabel 9. Potongan hasil perhitungan untuk *centroid* baru

NIS	Nama	P. Agama	PKn	B. Indo	B. Ing	MTK	IPA	IPS	S. Bud	Pengas	B. Jawa	Tata Busana	Cluster Baru				
													C1	C2	C3	C4	C5
4745	Adna Maksimal Hafizto	74	70	72	45	48	45	73	86	60	72	58	83,6	87,3255814	86,38461538	72,88461538	75,64912281
4746	Aisyah Mutiara C.	82	70	86	65	56	43	68	88	68	86	69	78,5	80,81395349	80,15384615	79,11538462	69,03508772
4747	Akhid Rijal F.	56	84	62	40	48	48	60	58	65	76	75	74,9	73,72093023	77,07692308	70,23076923	63,42103263
4748	Alfian Fajar Nur P.	80	89	75	60	60	48	75	46	64	78	71	65,6	62,1627907	50,53846154	47,19230769	40,35087719
4750	Annita Rahmawati	88	82	82	65	56	63	70	72	65	92	74	69,6	71,90697674	66,30769231	60,38461538	52,96491228
4751	Banu Nurcahyo	70	76	74	45	68	48	70	78	67	68	59	65,9	54,40697674	49,30769231	48,34615385	51,16666667
4754	Dwi Pangling	79	63	74	45	48	43	70	68	60	68	58	79	80,58139535	76,34615385	73,26923077	73,92982456
4757	Juli Iskandar	83	85	84	75	70	68	93	82	62	82	70	81,8	80,44186047	78,88461538	56,5	66,96491228
4759	Nukle Tantina Artha	78	66	90	50	64	53	83	66	62	76	56	58,8	62,81395349	63,88461538	62,69230769	62,10526316
4761	Panji Eka Sagita	73	81	41	35	52	40	80	66	55	86	60	79,5	86,41860465	78,03846154	78,34615385	75,9122807
4762	Puput Dwi Wijayanti	78	93	72	65	68	45	85	74	60	90	71	68,2	70,39534884	65,65384615	64,5	62,29824561

6) Kembali ke langkah 3 jika posisi *centroid* baru dengan *centroid* lama tidak sama.

Langkah-langkah diatas memperoleh hasil posisi *centroid* baru dengan lama sama pada iterasi ke-18. Berikut hasil pengelompokkan data iterasi ke-15 ditunjukkan pada tabel 10.

Tabel 10. Potongan hasil pengelompokkan data iterasi ke-15

C1	C2	C3	C4	C5
		1		
		1		
	1			1
1			1	
		1		
1				1
		1		
				1
	1			
	1			
	1			
	1			

				1
	1			
		1		
		1		
1				
				2
1				
1				
1				
1				
		1		
			1	
		1		
			1	
				1

### 3.2 Pembahasan Hasil Metode Manual Dengan Metode Algoritma *K-Means Clustering*

SMP N 02 Tasikmadu dalam pemilihan kelas favorit 8A sampai 8D menggunakan metode pemilihan sesuai rangking yaitu dengan pemilihan kelas 8A untuk siswa rangking 1-5, 8B rangking 6-10, 8C rangking 11-15, 8D rangking 16-20, 8E rangking 21-24, 8F rangking 25-28, 8G rangking 29-34. Dalam metode ini masih kurang strategis dalam penentuan kelas unggulan sesuai nilai kemampuan siswa karena hanya berpacu pada rangking saja. Siswa unggulan yang terpilih dalam kelas favorit masih terlihat secara global. Sehingga metode ini kurang strategis dalam pemilihan siswa kelas favorit serta dalam memberikan rencana dalam peningkatan proses belajar seperti penambahan jam pelajaran bagi kelas reguler, jam tambahan bagi siswa kelas favorit. Berikut hasil metode rangking yang sudah dilaksanakan di sekolah SMP N 02 Tasikmadu yang seperti di tabel 11:

Tabel 11. Potongan hasil metode rangking kelas 8B

No.	NIS	Nama	P. Agama	PKn	B. Indo	B. Ing	MTK	IPA	IPS	S. Bud	Penjas	TIK	B. Jawa	Tata Busana
1	4746	Affiah	96	79	74	50	72	58	83	70	61	84	76	61
2	4762	Agung Sri G.	87	68	77	60	72	58	85	72	63	95	86	80
3	4764	Aisyah Mutiara C.	82	70	86	65	56	43	68	88	68	80	86	69
4	4773	Alsri Wahono	80	90	74	40	68	58	80	40	61	80	80	57
5	4775	Andini Catur Utami	81	84	77	38	52	62,5	75	72	70	80	88	74
6	4787	Andriyan	76	73	66	40	62	58	85	64	58	68	84	65
7	4788	Annisa Dita Kirana	74	76	82	45	62	60	80	70	61	64	78	69
8	4802	Annisa Nur Aini	98	89	70	50	78	55	80	98	65	94	88	63
9	4805	Azzahra Shalfadya Y	84	79	62	65	66	63	63	66	60	85	74	65
10	4806	Bimo Murdiyono	80	84	72	45	68	67,5	85	56	65	66	82	77

Dalam penerapan algoritma *k-means clustering* menghasilkan pengelompokan siswa kelas unggulan/favorit yang ditunjukkan pada tabel 12.

Tabel 12. Hasil *k-means clustering*

<i>Cluster 1</i>	<i>Cluster 2</i>	<i>Cluster 3</i>
<p>Kelas 8A :</p> <p>Dari kelas 7A ada 6 siswa dengan rangking 1,2,3,5,6,7.</p> <p>Dari kelas 7B ada 4 siswa dengan rangking 1,2,3,4.</p> <p>Dari kelas 7C ada 2 siswa dengan rangking 1,3.</p> <p>Dari kelas 7D ada 7 siswa dengan rangking 1,2,3,4,5,9,11.</p> <p>Dari kelas 7E ada 3 siswa dengan rangking 1,2,3.</p> <p>Dari kelas 7F ada 5 siswa dengan rangking 1,2,3,4,6.</p> <p>Dari kelas 7G ada 3 siswa dengan rangking 1,2,3.</p>	<p>KELAS 8B :</p> <p>Dari kelas 7A ada 6 siswa dengan rangking 4,8,9,11,14,16.</p> <p>Dari kelas 7B ada 6 siswa dengan rangking 5,6,7,8,9,10.</p> <p>Dari kelas 7C ada 4 siswa dengan rangking 2,4,5,8.</p> <p>Dari kelas 7D ada 5 siswa dengan rangking 6,7,8,14,15.</p> <p>Dari kelas 7 E ada 3 siswa dengan rangking 4,5,8.</p> <p>Dari kelas 7F ada 8 siswa dengan rangking 5,7,8,9,10,11,12,13.</p> <p>Dari kelas 7G ada 4 siswa dengan rangking 4,5,6,7.</p>	<p>KELAS 8C :</p> <p>Dari kelas 7A ada 6 siswa dengan rangking 10,11,13,17,18,21.</p> <p>Dari kelas 7B ada 6 siswa dengan rangking 11,12,13,14,18,21.</p> <p>Dari kelas 7C ada 7 siswa dengan rangking 6,7,9,10,15,17,20.</p> <p>Dari kelas 7D ada 7 siswa dengan rangking 10,12,13,16,17,19,20.</p> <p>Dari kelas 7E ada 2 siswa dengan rangking 7,14.</p> <p>Dari kelas 7F ada 2 siswa dengan rangking 15,18.</p>

Dari hasil penggunaan metode algoritma *k-means clustering* didapatkan bahwa karakteristik siswa pada *cluster* ke-1 didominasi oleh siswa yang berasal dari kelas 7A dan 7D. *Cluster* ke-2 didominasi oleh siswa yang berasal dari kelas 7A, 7B dan 7F. *Cluster* ke-3 didominasi oleh siswa yang berasal dari kelas 7C dan 7D.

#### 4. PENUTUP

*K-means clustering* merupakan metode klasterisasi berdasarkan persamaan karakteristik, dan merupakan metode yang sangat berguna karena mampu mentranslasi ukuran kuantitatif. Berdasarkan hasil pengelompokan data menggunakan algoritma *k-means clustering*, di dapatkan hasil *clustering* hingga iterasi ke-15, dimana titik pusat tidak lagi berubah dan tidak ada data yang berpindah antar *cluster*. Penelitian ini menggunakan 226 data uji, dimana ke-226 data tersebut dibagi menjadi 5 *cluster* dengan keterangan *cluster* ke-1 merupakan kelas favorit 1 (kelas 8A), *cluster* ke-2 kelas favorit 2 (kelas 8B), dan *cluster* 3 kelas favorit 3 (kelas 8C), *cluster* ke-4 sampai cluster ke-5 dan data siswa sisanya kelas reguler. Sehingga dibandingkan dari metode yang sudah ada hasil penelitian ini lebih membantu dalam pemilihan siswa kelas unggulan/favorit serta memberikan rencana strategis dalam proses pembelajaran seperti penambahan jam pelajaran bagi kelas reguler, jam tambahan bagi siswa kelas favorit. Sehingga dapat tercipta peningkatan proses belajar siswa untuk mencapai hasil optimal.

#### DAFTAR PUSTAKA

- Abdillah, L. A.. "*Students learning center strategy based on e-learning and blogs*," in Seminar Nasional Sains dan Teknologi (SNST) ke-4 Tahun 2013, Fakultas Teknik Universitas Wahid Hasyim Semarang 2013, pp. F.3.15-20.
- Aggarwal, Charu C. & Reddy, Chandan K.(Ed.). 2013. *Data Clustering: Algorithms and Applications*. Boca Raton : CRC Press.
- Bhardwaj, Ankit. Sharma, Arvind. Shrivastava, V.K.. (2012). "*Data Mining Techniques and Their Implementation in Blood BankSector - A Review*". *International Journal of Engineering Research and Applications (IJERA)* ISSN: 2248- 9622, Vol. 2, Issue4, July-August 2012, pp.1303-1309.
- Hermawati, Fajar Astuti. 2013. *Data Mining*. Yogyakarta: Andi.
- Hastuti, Khafiizh. (2012). *Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Mahasiswa Non Aktif*. Semarang : Seminar Nasional Teknologi Informasi & Komunikasi Terapan 2012. ISBN 979 - 26 - 0255 - 0.
- Prasetyo, Eko. 2012. *Data Mining - Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Penerbit Andi.

Satrio Wahono, Romi. (2012). *Proses Data Mining*. From : [http://www. Romisatria wahono.net/lecture/dm/romi-dm-02-proses-june2012.pptx](http://www.Romisatria wahono.net/lecture/dm/romi-dm-02-proses-june2012.pptx) , diakses tanggal 16 Desember 2013 jam 15.30 WIB.

Sujarweni, V Wiratna. 2014. *Metodologi Penelitian*. Yogyakarta: PUSTAKABARUPERSS.

Untari, Dwi. 2014. *Data Mining Untuk Menganalisa Prediksi Mahasiswa Berpotensi Non-Aktif Menggunakan Metode decision Tree C4.5*. skripsi. Semarang: Universitas Dian Nuswantoro.